



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|-----------------|-------------|----------------------|---------------------|------------------|
|-----------------|-------------|----------------------|---------------------|------------------|

10/706,282

11/13/2003

Aviv Bachar

0004794USU/4269

5641

27623

7590

09/28/2010

OHLANDT, GREELEY, RUGGIERO & PERLE, LLP  
ONE LANDMARK SQUARE, 10TH FLOOR  
STAMFORD, CT 06901

EXAMINER

COLUCCI, MICHAEL C

ART UNIT

PAPER NUMBER

2626

MAIL DATE

DELIVERY MODE

09/28/2010

PAPER

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

|                              |                                       |                                      |  |
|------------------------------|---------------------------------------|--------------------------------------|--|
| <b>Office Action Summary</b> | <b>Application No.</b><br>10/706,282  | <b>Applicant(s)</b><br>BACHAR ET AL. |  |
|                              | <b>Examiner</b><br>MICHAEL C. COLUCCI | <b>Art Unit</b><br>2626              |  |

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

### Status

- 1) ☒ Responsive to communication(s) filed on 26 July 2010.
- 2a) ☒ This action is **FINAL**.                      2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

### Disposition of Claims

- 4) ☒ Claim(s) 1-3,9,10,12,15-17,19,21,23,25-30,32-39,41-43,46 and 47 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-3,9,10,12,15-17,19,21,23,25-30,32-39,41-43,46 and 47 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

### Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

### Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All    b) ☐ Some \*    c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

### Attachment(s)

- |  |   |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)          | 4) <input type="checkbox"/> Interview Summary (PTO-413)           |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date. _____                                      |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)          | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| Paper No(s)/Mail Date _____  | 6) <input type="checkbox"/> Other: _____                          |

## DETAILED ACTION

### *Response to Arguments*

1. Applicant's arguments filed 02/22/2010 have been fully considered but they are not persuasive.

#### **Argument (page 9):**

- “Neither Scarano nor Walsh discloses that the analysis types are selected such that the second analysis type requires more computing resources than the first audio analysis type”

#### **Response to argument:**

Examiner disagrees, consider that while giving claims their broadest reasonable interpretation in light of the supporting disclosure without importing limitations from the specification into the claims unnecessarily, Scarano initially teaches a first analysis type, when a new audio segment is available a decision is made at step 1904 whether that audio should be processed. If there is no CTI data some information may be provided by the recording device at 1902 such as which phone extension or trunk provided the audio. If the optional CTI interface is included, there is additional data as noted in connection with 1903. Using all available data logic is executed at 1904 and a decision is made about the audio segment. If the decision is to process the audio, then a reference to the audio and it's associated data is put in a queue for speech processing ([0159]).

Second, Scarano teaches another analysis type, processing 1905 is alerted when a reference to an audio segment is added to the queue, it invokes the speech engine to pre process the audio into an intermediate format. The intermediate format is a representation of the audio that is optimized for rapid searching. Some representations that are suitable for rapid searches are a statistical model of the phonemes or a text representation of the contents of the audio. Once the intermediate format is created, then rules determination is executed at 1906 ([0160]).

Examiner believes that Walsh improves the resource management of Scarano by implementing processors to handle calls in a system by managing conferences within an audio conferencing system, the method comprising: identifying a first resource with a predetermined capacity to receive additional conferences, the first resource having a plurality of channels and operating under control of a processor to handle audio conferences; identifying a second resource with a predetermined capacity to receive additional conferences, the second resource having a plurality of channels and operating under control of a processor to handle audio conferences, the capacity of the second resource being less than the capacity of the first resource, and the second resource including a conference; moving the conference on the second resource to the first resource if the first resource has a capacity to include the conference, and attempting to identify a third resource if the first resource does not have the capacity to include the conference; for respective conferences (Walsh Col. 12 lines 15-30).

Therefore, Walsh improves Scarano by implementing a first and second resource based decision, for example, using channels and resource decision making to improve the call handling of Scarano, such as by alleviating past difficulties

that occur in a free seating environment, wherein it may be difficult to directly identify and locate calls for a particular CSR. Further, recorded call segments that span conferences and transfers may be difficult to accurately be accounted for (Scarano [0063]).

Therefore, Walsh can shift resources during the decision making of Scarano to accurately account for conference calls or individual calls where the capacity of the second resource is less than the capacity of the first resource, and the second resource may include a conference, wherein moving the conference on the second resource to the first resource if the first resource has a capacity to include the conference (Walsh Col. 12 lines 15-30).

**Argument (page 9):**

- “Neither Scarano nor Walsh discloses first and second audio analysis components”

**Response to argument:**

Examiner disagrees, wherein an audio analysis component can be virtually any component which process audio. Figure 19 of Scarano show a first and second audio analysis component such as for example elements 1904 and 1905.

Art Unit: 2626

Consider that giving claims their broadest reasonable interpretation in light of the supporting disclosure without importing limitations from the specification into the claims unnecessarily, Scarano initially teaches a first analysis type, when a new audio segment is available a decision is made at step 1904 whether that audio should be processed. If there is no CTI data some information may be provided by the recording device at 1902 such as which phone extension or trunk provided the audio. If the optional CTI interface is included, there is additional data as noted in connection with 1903. Using all available data logic is executed at 1904 and a decision is made about the audio segment. If the decision is to process the audio, then a reference to the audio and it's associated data is put in a queue for speech processing ([0159]).

Second, Scarano teaches another analysis type, processing 1905 is alerted when a reference to an audio segment is added to the queue, it invokes the speech engine to pre process the audio into an intermediate format. The intermediate format is a representation of the audio that is optimized for rapid searching. Some representations that are suitable for rapid searches are a statistical model of the phonemes or a text representation of the contents of the audio. Once the intermediate format is created, then rules determination is executed at 1906 ([0160]).

These are explicitly two analysis types, wherein the second component (processing and format selection) may require more resources than the first (yes or no decision). Examiner believes that Walsh renders

**Argument (page 10):**

- “Neither Scarano nor Walsh discloses that the first audio analysis is activated on a part of the interaction that is an initial region of interest of an interaction”

**Response to argument:**

Examiner disagrees, wherein Scarano teaches audio segments extracted from a conversation. If only a segment of audio is extracted from a larger portion of audio, this segment is a region of interest ([0157] [0159]).

Further, Scarano explicitly teaches analysis on a “part of the interaction”, wherein Scarano teaches the process of threshold determination is performed by first determining a set of calls that represent a test or training set. A specific phrase is selected, a search is performed, and the resulting list of result hypotheses will be returned. A human listener is then used to listen to the list of result hypotheses and to determine at what point in the result distribution that the confidence scores fail to be accurate. As the listener inspects search results, they are queued to the exact point in each call that the candidate result was located and allows the listener to only listen to a small portion of each call in order to determine the appropriate threshold (Scarano [0166]).

**Argument (page 11):**

- “Neither Scarano nor Walsh discloses that the second audio analysis is activated on a part of the interaction that is a region of interest of an interaction”

**Response to argument:**

Examiner disagrees, wherein Scarano teaches alerting audio analysis component 1905 in Figure 19 to process an audio segment ([0160]).

Further, Scarano explicitly teaches analysis on a “part of the interaction”, wherein Scarano teaches the process of threshold determination is performed by first determining a set of calls that represent a test or training set. A specific phrase is selected, a search is performed, and the resulting list of result hypotheses will be returned. A human listener is then used to listen to the list of result hypotheses and to determine at what point in the result distribution that the confidence scores fail to be accurate. As the listener inspects search results, they are queued to the exact point in each call that the candidate result was located and allows the listener to only listen to a small portion of each call in order to determine the appropriate threshold (Scarano [0166]).

**Argument (page 11):**

- “Neither Scarano nor Brown discloses activating the first audio analysis component for dynamically reducing the initial region of interest to obtain the region of interest”

**Response to argument:**

Examiner disagrees, wherein Scarano teaches searching a set of audio segment for the phrase; and producing a set of results of all occurrences of the phrase within the audio segments. In other words, an audio segment is extracted from a conversation, and the audio segment is further narrowed to find a specific phrase ([0030]).

Further, Scarano explicitly teaches analysis on a “part of the interaction”, wherein Scarano teaches the process of threshold determination is performed by first determining a set of calls that represent a test or training set. A specific phrase is selected, a search is performed, and the resulting list of result hypotheses will be returned. A human listener is then used to listen to the list of result hypotheses and to determine at what point in the result distribution that the confidence scores fail to be accurate. As the listener inspects search results, they are queued to the exact point in each call that the candidate result was located and allows the listener to only listen to a small portion of each call in order to determine the appropriate threshold (Scarano [0166]).

**Argument (page 12):**

- “In forming the section 103 rejection, the Office Action acknowledges that Scarano in view of Walsh fails to teach the usage of screen events, and introduces Bscheider for this proposition”

**Response to argument:**

Examiner disagrees, wherein Bscheider teaches a GUI for call handling, wherein the close of a play session (e.g., the user hits Stop or Pause in a typical audio playback

Art Unit: 2626

GUI displayed in conjunction with the GUI described in FIG. 16) a StopStream call is made to the SCM. The thread in turn detects that the stopped state has been entered, exits from the request loop code, and frees up any used resources. Finally, it informs the client that a Stop event has occurred. If the entire call record is played without calling StopStream, the SCM performs the same exit and cleanup code, but informs the client that a Done event has occurred instead (Bscheider Col. 56 liners 42-51 & Fig. 16).

**Argument (page 13):**

- “In forming me §103 rejection, the Office Action acknowledges that Scarano in view of Walsh fails to teach that the method is used for verifying that an agent requested a customer's permission to put the customer on hold, wherein the pivot spot is the time the agent put the customer on hold, the initial region of interest is the whole interaction, and wherein the region of interest is defined by a first predetermined number of seconds prior to the pivot spot and a second predetermined number of seconds following the hold”

**Response to argument:**

Examiner disagrees, wherein Eilbacher teaches a contact center 200 of FIG. 2, and in particular a telephone call center. Referring to FIG. 3, customers 100 access the contact center through the public switched telephone network (PSTN) 101 and an automatic call distribution system 102 (PBX/ACD) directs the communication to one of a plurality of agent work stations 104. Each agent work station 104 includes, for example,

Art Unit: 2626

a computer and a telephone set. Communications are directed to the agent stations 104 based on the availability of the agent. In those contact centers handling communications for a number of different clients, communications to a particular client may be routed to a finite group of agents specifically trained to respond to the needs of that customer or that client. Alternatively, the PBX/ACD 102 may include an interactive voice response (IVR) system that presents an audio menu to a customer, requesting a response by way of the customer's telephone key pad or by way of a voice response. Then, a call is directed to a particular group of agent stations 104 or to a particular information retrieval system, based on the responses of the customer. For example, the system can provide the customer 100 with the address to which products should be returned or the Internet address for obtaining additional product information. All data associated with the customer's communication and the agents responsive interaction with the customer may be recorded by a monitor module 210 within monitoring system 204. Examples of the data typically recorded by a telephone call center system include the audio communication between the customer and the agent, key pad data input by the customer, screens viewed by the agent on the computer at the agent station 104 (carried by data line 105), the start and end time for the customer's communication, the identity of the customer, including the originating telephone number and the called number, the identity of the various agents servicing the communications, the length of time the customer is on hold and the steps the customer navigated before terminating the communication (Eilbacher Col. 8 lines 29-67).

Art Unit: 2626

Further, Eilbacher teaches incoming and outgoing calls can be recorded in their entirety; particular calls can be identified for recording, such as by client or agent; and calls can be recorded by event, such as calls exceeding five minutes. If "cradle-to-grave" recording is used, then all information related to a particular telephone call or caller-initiated transaction is recorded, from the time the call enters the contact center to the later of: the caller hanging up or the agent completing the transaction. All of the interactions during the call are recorded, including interaction with the IVR system, time spent on hold, data keyed through the caller's key pad, conversations with the agent, and screens displayed by the agent at his/her station 104 during the transaction. These types of recordings allow for evaluation of the full customer experience throughout the transaction. As an example, the length of time a customer was on hold during a purchase transaction can be analyzed as a possible deterrent to completing a purchase. Such information may be used by contact center managers to modify their procedures, staffing, and/or equipment to improve the customer's experience when using the contact center. The comprehensiveness of the data capture of the present invention also allows for the subsequent verification of transaction content. For example, a dispute over what information was verbally provided by a caller applying for insurance coverage over the telephone can easily be resolved by replaying the application call in its entirety. Whether a customer selected size 10 can also be proven, as can whether the customer/investor authorized the purchase of 100 shares of a particular stock. (Eilbacher Col. 9 lines 10-39).

Art Unit: 2626

Furthermore, Eilbacher teaches types of parameters which can be analyzed by the customer experience analyzing unit 208 include the number of key strokes entered by the customer during a telephone call, the length of a telephone call, time on hold, number of transfers, or length of a queue. That is, if the length of the telephone call, the number of key strokes entered during the call or the length of a queue exceeded predetermined levels, the customer experience analyzing unit 208 can determine that the communication was likely unsatisfactory. In addition, speech detection or word spotting can be used to detect certain inflammatory words such as curse words. For example, in the case of word spotting, an audio analysis is performed on recorded audio such as a telephone call. The audio is automatically processed, searching for any key words on a predefined list which have been identified as cause for concern. If any of the words are found, the call is marked as a potentially negative customer experience. This word spotting audio analysis can be done separately, or in addition to the stress audio analysis. Similarly, in connection with an e-mail communication, a text search can be used to look for words such as curse words, which might tend to indicate an unsatisfactory customer experience (Eilbacher Col. 11 lines 25-61).

Therefore, Eilbacher improves Scarano that is

“verifying that an agent requested a customer's permission to put the customer on hold, wherein the pivot spot is the time the agent put the customer on hold, the initial region of interest is the whole interaction, and wherein the region of interest is defined by a first predetermined number of seconds prior to the pivot spot and a second predetermined number of seconds following the hold”

for instance by taking into account the length of a queue exceeded predetermined levels, the customer experience analyzing unit 208 can determine that the communication was likely unsatisfactory where speech detection or word spotting can be used to detect certain inflammatory words such as curse words (Eilbacher Col. 11 lines 25-61).

**Argument (page 14):**

- “In forming the §103 rejection, the Office Action acknowledges that Scarano in view of Walsh fails to teach that the method is used for measuring the effectiveness of a promotion offer to a customer requesting the termination of the service, wherein the pivot spot is the time of a screen event related to offering a promotion or to an account being saved or lost, and wherein the region of interest is defined by a first predetermined number of seconds prior to the pivot spot.”

**Response to argument:**

Art Unit: 2626

Examiner disagrees, wherein Bernard explicitly teaches that promotional items are offered to a customer 182 based on his or her calling and purchasing history. For example, in one embodiment, the automated product purchasing system review calling and purchasing statistics maintained for a shopper 182. Statistics can be maintained by VRU 104, interactive transaction database 112, or even by reporting database 438. If these statistics indicate that the shopper is a particularly good customer of the automated product purchasing system, interface unit 104 may offer a promotional or special item to that shopper 182. For example, where shopper 182 is a frequent purchaser, interface unit 104 may inform him or her that upon the next purchase, he or she will receive a bonus CD (Bernard Col. 51 lines 29-41).

Art Unit: 2626

Though, Eilbacher teaches identifying time periods for instance by taking into account the length of a queue exceeded predetermined levels, the customer experience analyzing unit 208 can determine that the communication was likely unsatisfactory where speech detection or word spotting can be used to detect certain inflammatory words such as curse words (Eilbacher Col. 11 lines 25-61), Bernard improves Eilbacher and teaches the evaluation of unsatisfactory or satisfactory calls such as by a customer rating by analyzing usefulness of promotions or special items.

### ***Claim Rejections - 35 USC § 103***

2. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

3. Claims 1-3, 9, 10, 15-17, 19, 21, 23, 25-27, 29, 32-39, 41, and 46-47 are rejected under 35 U.S.C. 103(a) as being unpatentable over Scarano et al. US 20040083099 A1 (hereinafter Scarano) in view of Walsh et al. US 6539087 B1 (hereinafter Walsh).

Re claims 1 and 19, Scarano teaches an apparatus for event-driven content analysis of an audio captured interaction captured in a call center, within a

Art Unit: 2626

computerized system having a processing unit and a storage unit ([0161]), the apparatus comprising the elements of:

- an audio or video recording device for recording the audio interaction ([0162]);

- a pivot spot defining component for automatically marking an at least one time position in the audio interaction that indicates the occurrence of an at least one pre-defined event or data item ([0030]);

- a first audio analysis component of a first audio analysis type (Fig. 19 item 1902);

- a region of interest defining component for defining an initial region of interest, by determining the time limits of an at least one segment of the audio interaction ([0108]), the segment containing the time position of a pivot spot ([0115]),

- and for activating the first audio analysis component for dynamically reducing the time ([0010]) limits of the initial region of interest to obtain the region of interest ([0098-0102], phrase isolation);

- a second audio analysis component of a second audio analysis type for analyzing the region of interest of the audio interaction ([0098-0102] & Fig. 19 items 1906-1908),

However, Scarano fails to teach the first audio analysis type and the second audio analysis type are selected such that the second audio analysis type requires more computing resources than the first audio analysis type.

Walsh teaches managing conferences within an audio conferencing system, the method comprising: identifying a first resource with a predetermined capacity to receive

Art Unit: 2626

additional conferences, the first resource having a plurality of channels and operating under control of a processor to handle audio conferences; identifying a second resource with a predetermined capacity to receive additional conferences, the second resource having a plurality of channels and operating under control of a processor to handle audio conferences, the capacity of the second resource being less than the capacity of the first resource, and the second resource including a conference; moving the conference on the second resource to the first resource if the first resource has a capacity to include the conference, and attempting to identify a third resource if the first resource does not have the capacity to include the conference; for respective conferences (Walsh Col. 12 lines 15-30).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Scarano to incorporate the first audio analysis type and the second audio analysis type are selected such that the second audio analysis type requires more computing resources than the first audio analysis type as taught by Walsh to allow for a conferencing system that dynamically assigns calls to audio processing resources, wherein the system may attempt to process each audio conference on a single audio processing resource, so that information about conference participants does not need to be shared across audio processing resources (Walsh Col. 1 lines 35-45), where a shift in resources during the decision making of Scarano would accurately account for conference calls or individual calls where the capacity of the second resource is less than the capacity of the first resource, and the second resource may include a conference, wherein moving the conference on the

Art Unit: 2626

second resource to the first resource if the first resource has a capacity to include the conference (Walsh Col. 12 lines 15-30).

Re claim 2, Scarano teaches the apparatus of claim 1 further comprising a content audio analysis input selector component to determine an at least one input or parameter for an at least one analyzer component ([0115])

Re claim 3, Scarano teaches the apparatus of claim 1 further comprises an audio analysis type selector component to identify and to select an at least one analyzer component type for determining the Region of Interest ([0098-0102].

Re claim 9, Scarano teaches the apparatus of claim 1 further comprises the element of an audio analyzer component to analyze the audio elements of the interaction data ([0108])

Re claim 10, Scarano teaches the apparatus of claim 1 wherein the first audio analysis component of the second audio analysis component is a computer telephony interface events analyzer component for analyzing at least one common telephony events associated with the interaction data ([0006]).

Art Unit: 2626

Re claim 15, Scarano teaches the apparatus of claim 14 wherein the interaction media is at least one data packet carrying voice or other media over internet protocol ([0156]).

Re claim 16, Scarano teaches the apparatus of claim 1 wherein the region of interest is a specific segment of the interaction media that is analyzed to extract meaningful interaction-specific information in an organization ([0108])

Re claim 17, Scarano teaches the apparatus of claim 1 wherein the interaction associated with an at least one computer telephony integration event occurring during the interaction ([0006]).

Re claim 21, Scarano teaches the method of claim 19 further comprising the step of selecting a method for the audio analysis of the at least one interaction media based on the at least one event associated with the interaction ([0108]).

Re claim 23, Scarano teaches the method of claim 19 further comprising the step of selecting the parameters to be used in the at least one audio analysis instruction step on the at least one segment of the interaction ([0098-0102] & Fig. 19 items 1906-1908).

Re claim 25, Scarano teaches the method of claim 19 wherein the region of interest is predetermined by a user or an apparatus ([0115])

Re claim 26, Scarano teaches the method of claim 19 further comprises the steps of receiving interaction data and associated meta-data from an at least one interaction ([0078]).

Re claim 27, Scarano teaches the method of claim 19 wherein the at least one first audio analysis or the second audio analysis instruction step comprises the step of analyzing the-speech elements of the interaction data for the presence of pre-defined words or phrases ([0108]).

Re claim 29, Scarano teaches the method of claim 19 wherein the at least one first audio analysis instruction step or the second audio analysis comprises the steps of analyzing t4ae-speech elements of the interaction data for pre-defined speech patterns ([0108]).

Re claim 32, Scarano teaches the method of claim 19 further comprises performing an at least one content audio analysis step during capturing of the interaction data and the interaction meta-data ([0065], start and end times embedded).

Re claim 33, Scarano teaches the method of claim 19 wherein the at least one pivot spot or the region of interest ([0115]) are determined based on an event external to the interaction ([0055], separate CTI actions).

Re claims 34 and 35, Scarano teaches the apparatus of claim 1 wherein the pivot spot is determined using at least one item selected from the group consisting of: a Computer Telephony Integration event ([0065]); a screen event; an emotional level; and a spotted word.

Re claims 36 and 37, Scarano teaches the apparatus of claim 1 wherein the first audio analysis component used for reducing the initial region of interest is selected from the group consisting of: an emotional level audio analysis component, a word spotting audio analysis component ([0067] search for spoken words), audio event analysis component, dual tone multi frequency (DTMF) event audio analysis component, and even priority audio analysis component.

Re claims 38 and 39, Scarano teaches the apparatus of claim 1 wherein the captured interaction is between an agent and a customer ([0028], well known call center operations between caller/customer and agent).

Re claim 41, Scarano teaches the method of claim 19 wherein the method is used for detecting customer churn indications ([0098-0102], cancel service), wherein the pivot spot is defined using a CTI hold event or a cancellation-related screen event;

Art Unit: 2626

and wherein the region of interest is defined using emotion audio analysis or word spotting ([0065], audio segments).

Re claim 46, Scarano teaches the apparatus of claim 1 wherein the at least one pivot spot or the region of interest are determined based on an event external to the interaction ([0055], separate CTI actions).

Re claim 47, Scarano teaches the method of claim 19 wherein the reducing step is repeated two or more times ([0036], more than one audio segment from multiple extraction operations).

**4. Claim 30 is rejected under 35 U.S.C. 103(a) as being unpatentable over Scarano et al. US 20040083099 A1 (hereinafter Scarano) in view of Walsh et al. US 6539087 B1 (hereinafter Walsh) and further in view of Bscheider et al. US 6937706 B2 (hereinafter Bscheider).**

Re claim 30, Scarano teaches the method of claim 19 further comprises the steps of:

identifying an at least one pre-defined computer telephony integrated event in the interaction data ([0108]);

However, Scarano in view of Walsh fails to teach identifying an at least one pre-defined screen event in the interaction data

Bscheider teaches the close of a play session (e.g., the user hits Stop or Pause in a typical audio playback GUI displayed in conjunction with the GUI described in FIG. 16) a StopStream call is made to the SCM. The thread in turn detects that the stopped state has been entered, exits from the request loop code, and frees up any used resources. Finally, it informs the client that a Stop event has occurred. If the entire call record is played without calling StopStream, the SCM performs the same exit and cleanup code, but informs the client that a Done event has occurred instead (Bscheider Col. 56 liners 42-51 & Fig. 16).

Further, Bscheider teaches that the term "Call Control" refers to the part of the metadata concerning the creation and termination of call records. The term "Media" refers to the actual data that is being recorded. This term is used interchangeably with audio since the primary design of the CRG is to support audio recording. However, the CRG could apply to any data being recorded including multimedia or screen image data (Col. 31 lines 29-40).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Scarano in view of Walsh to incorporate identifying an at least one pre-defined screen event in the interaction data as taught by Bscheider to allow for the display of events on a user screen, wherein a user can view interactions of one or more agents and the time periods of interaction during a call (Bscheider Col. 56 liners 42-51 & Fig. 16).

**5. Claims 12 and 28 are rejected under 35 U.S.C. 103(a) as being unpatentable over Scarano et al. US 20040083099 A1 (hereinafter Scarano) in view of Walsh et al. US 6539087 B1 (hereinafter Walsh) and further in view of Petrushin US 20020194002 A1 (hereinafter Petrushin).**

Re claims 12, Scarano teaches the apparatus of claim 9, wherein the audio analyzer component further comprises the elements of:

a word spotting component to locate and identify pre-defined terms or patterns in the speech elements of the interaction data ([0115]);

a talk analyzer component to identify and locate specific pre-defined speech events in the speech elements of the information data ([0108])

However, Scarano in view of Walsh fails to teach an emotion audio analysis component to locate and identify positive or negative emotions in the interaction data

Petrushin teaches logic for receiving and analyzing a speech signal, logic for dividing the speech signal, and logic for extracting at least one feature from the speech signal. The system comprises logic for calculating statistics of the speech, and logic for at least one neural network for classifying the speech as belonging to at least one of a finite number of emotional states. The system also comprises logic for outputting an indication of the at least one emotional state (Petrushin [0011]).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Scarano in view of Walsh to incorporate an emotion audio analysis component to locate and identify positive or negative emotions in the interaction data as taught by Petrushin to allow for the classification of a speech

Art Unit: 2626

based on emotion, wherein features from a signal are used to determine an emotional state (Petrushin [0011]).

Re claim 28, Scarano teaches the method of claim 19 wherein the at least one first audio analysis instruction step or the second audio analysis comprises the step of analyzing the-speech elements of the interaction data ([0115]) to detect positive and negative emotions.

However, Scarano in view of Walsh fails to teach analyzing the-speech elements of the interaction data ([0115]) to detect positive and negative emotions.

Petrushin teaches logic for receiving and analyzing a speech signal, logic for dividing the speech signal, and logic for extracting at least one feature from the speech signal. The system comprises logic for calculating statistics of the speech, and logic for at least one neural network for classifying the speech as belonging to at least one of a finite number of emotional states. The system also comprises logic for outputting an indication of the at least one emotional state (Petrushin [0011]).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Scarano in view of Walsh to incorporate analyzing the-speech elements of the interaction data ([0115]) to detect positive and negative emotions as taught by Petrushin to allow for the classification of a speech based on emotion, wherein features from a signal are used to determine an emotional state (Petrushin [0011]).

**6. Claim 42 rejected under 35 U.S.C. 103(a) as being unpatentable over Scarano et al. US 20040083099 A1 (hereinafter Scarano) in view of Walsh et al. US 6539087 B1 (hereinafter Walsh) and further in view of Eilbacher et al. US 6724887 B1 (hereinafter Eilbacher).**

Re claim 42, Scarano teaches the method of claim 19 wherein the method is used for verifying that an agent requested a customer's permission to put the customer on hold, wherein the pivot spot ([0030]) is the time the agent put the customer on hold, the initial region of interest is the whole interaction ([0010]), and wherein the region of interest is defined by a first predetermined number of seconds prior to the pivot spot and a second predetermined number of seconds following the hold.

However, Scarano in view of Walsh fails to teach the region of interest is defined by a first predetermined number of seconds prior to the pivot spot and a second predetermined number of seconds following the hold

Eilbacher teaches a contact center 200 of FIG. 2, and in particular a telephone call center. Referring to FIG. 3, customers 100 access the contact center through the public switched telephone network (PSTN) 101 and an automatic call distribution system 102 (PBX/ACD) directs the communication to one of a plurality of agent work stations 104. Each agent work station 104 includes, for example, a computer and a telephone set. Communications are directed to the agent stations 104 based on the availability of the agent. In those contact centers handling communications for a number of different clients, communications to a particular client may be routed to a

Art Unit: 2626

finite group of agents specifically trained to respond to the needs of that customer or that client. Alternatively, the PBX/ACD 102 may include an interactive voice response (IVR) system that presents an audio menu to a customer, requesting a response by way of the customer's telephone key pad or by way of a voice response. Then, a call is directed to a particular group of agent stations 104 or to a particular information retrieval system, based on the responses of the customer. For example, the system can provide the customer 100 with the address to which products should be returned or the Internet address for obtaining additional product information. All data associated with the customer's communication and the agents responsive interaction with the customer may be recorded by a monitor module 210 within monitoring system 204. Examples of the data typically recorded by a telephone call center system include the audio communication between the customer and the agent, key pad data input by the customer, screens viewed by the agent on the computer at the agent station 104 (carried by data line 105), the start and end time for the customer's communication, the identity of the customer, including the originating telephone number and the called number, the identity of the various agents servicing the communications, the length of time the customer is on hold and the steps the customer navigated before terminating the communication (Eilbacher Col. 8 lines 29-67).

Further, Eilbacher teaches incoming and outgoing calls can be recorded in their entirety; particular calls can be identified for recording, such as by client or agent; and calls can be recorded by event, such as calls exceeding five minutes. If "cradle-to-grave" recording is used, then all information related to a particular telephone call or

Art Unit: 2626

caller-initiated transaction is recorded, from the time the call enters the contact center to the later of: the caller hanging up or the agent completing the transaction. All of the interactions during the call are recorded, including interaction with the IVR system, time spent on hold, data keyed through the caller's key pad, conversations with the agent, and screens displayed by the agent at his/her station 104 during the transaction. These types of recordings allow for evaluation of the full customer experience throughout the transaction. As an example, the length of time a customer was on hold during a purchase transaction can be analyzed as a possible deterrent to completing a purchase. Such information may be used by contact center managers to modify their procedures, staffing, and/or equipment to improve the customer's experience when using the contact center. The comprehensiveness of the data capture of the present invention also allows for the subsequent verification of transaction content. For example, a dispute over what information was verbally provided by a caller applying for insurance coverage over the telephone can easily be resolved by replaying the application call in its entirety. Whether a customer selected size 10 can also be proven, as can whether the customer/investor authorized the purchase of 100 shares of a particular stock. (Eilbacher Col. 9 lines 10-39).

Furthermore, Eilbacher teaches types of parameters which can be analyzed by the customer experience analyzing unit 208 include the number of key strokes entered by the customer during a telephone call, the length of a telephone call, time on hold, number of transfers, or length of a queue. That is, if the length of the telephone call, the number of key strokes entered during the call or the length of a queue exceeded

Art Unit: 2626

predetermined levels, the customer experience analyzing unit 208 can determine that the communication was likely unsatisfactory. In addition, speech detection or word spotting can be used to detect certain inflammatory words such as curse words. For example, in the case of word spotting, an audio analysis is performed on recorded audio such as a telephone call. The audio is automatically processed, searching for any key words on a predefined list which have been identified as cause for concern. If any of the words are found, the call is marked as a potentially negative customer experience. This word spotting audio analysis can be done separately, or in addition to the stress audio analysis. Similarly, in connection with an e-mail communication, a text search can be used to look for words such as curse words, which might tend to indicate an unsatisfactory customer experience (Eilbacher Col. 11 lines 25-61).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Scarano in view of Walsh to incorporate verifying that an agent requested a customer's permission to put the customer on hold, wherein the pivot spot is the time the agent put the customer on hold, the initial region of interest is the whole interaction, and wherein the region of interest is defined by a first predetermined number of seconds prior to the pivot spot and a second predetermined number of seconds following the hold as taught by Eilbacher to allow for the monitoring of a customer and agent interaction, wherein various retrieval methods are implemented based on responses during the wait time as well as during the actual interaction with an agent (Eilbacher Col. 6 lines 35-40) taking into account the length of a queue exceeded predetermined levels, the customer experience analyzing unit 208 can determine that

Art Unit: 2626

the communication was likely unsatisfactory where speech detection or word spotting can be used to detect certain inflammatory words such as curse words (Eilbacher Col. 11 lines 25-61).

**7. Claim 43 is rejected under 35 U.S.C. 103(a) as being unpatentable over Scarano et al. US 20040083099 A1 (hereinafter Scarano) in view of Eilbacher et al. US 6724887 B1 (hereinafter Eilbacher) and Walsh et al. US 6539087 B1 (hereinafter Walsh) and further in view of further in view of Bernard et al. US 5918213 A (hereinafter Bernard).**

Re claim 43, Scarano in view of Walsh and Eilbacher fails to teach the method of claim 19 wherein the method is used for measuring the effectiveness of a promotion offer to a customer requesting the termination of the service, wherein the pivot spot is the time of a screen event related to offering a promotion or to an account being saved or lost, and wherein the region of interest is defined by a first predetermined number of seconds prior to the pivot spot.

Bernard teaches that promotional items are offered to a customer 182 based on his or her calling and purchasing history. For example, in one embodiment, the automated product purchasing system review calling and purchasing statistics maintained for a shopper 182. Statistics can be maintained by VRU 104, interactive transaction database 112, or even by reporting database 438. If these statistics indicate that the shopper is a particularly good customer of the automated product purchasing

Art Unit: 2626

system, interface unit 104 may offer a promotional or special item to that shopper 182. For example, where shopper 182 is a frequent purchaser, interface unit 104 may inform him or her that upon the next purchase, he or she will receive a bonus CD (Bernard Col. 51 lines 29-41).

Further, Bernard teaches a caller 182 decides to delete all of the items in his or her virtual shopping cart, as indicated by input step 3724, VRU 104 simply deletes these items from the order. This deletion step is illustrated by a step 3728. In one embodiment this is accomplished by deleting all the order information from the previous call and beginning anew with fresh order information for the present call. In an alternative embodiment, the deletion is accomplished by simply removing the items' catalog ID numbers 1008 from the order information in interactive transaction database 112. A confirmation script can be played by VRU 104 announcing that the order has been canceled. As with the other options, at this time the caller is returned to the shopping mode where he or she can sample additional selections or terminate the phone call. Although not illustrated, in one embodiment, caller 182 is given the option of hearing a listing of the items in his or her virtual shopping cart before deciding whether to keep or cancel the order entirely. Finally, caller 182 may decide to individually review the items in the virtual shopping cart and determine whether each individual item is to be kept. If this is the case, caller 182 elects to review the items on hold as illustrated by input step 3732. In response, in a step 3736, VRU 104 reviews the order with caller 182. In one embodiment, this is accomplished by a process similar to that illustrated in FIG. 36 where each item is reviewed one at a time, and caller 182 selects whether to

Art Unit: 2626

accept or delete each item. Once the review order process is complete, caller 182 is forwarded to the shopping mode where he or she can sample additional selections, immediately purchase the selections remaining in his or her virtual shopping cart, or terminate the call (Bernard Col. 50 lines 1-32).

Therefore, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify the system of Scarano in view of Walsh and Eilbacher to incorporate measuring the effectiveness of a promotion offer to a customer requesting the termination of the service, wherein the pivot spot is the time of a screen event related to offering a promotion or to an account being saved or lost, and wherein the region of interest is defined by a first predetermined number of seconds prior to the pivot spot as taught by Bernard to allow for the automated and statistical determination of whether a promotional/coupon should be applied to a caller based on the callers history (Bernard Col. 51 lines 29-41), wherein an order can be cancelled based on a caller history (Bernard Col. 50 lines 1-32).

### ***Conclusion***

8. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within

Art Unit: 2626

TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to MICHAEL C. COLUCCI whose telephone number is (571)270-1847. The examiner can normally be reached on 8:30 am - 5:00 pm , Monday - Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571)-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Art Unit: 2626

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Michael C Colucci/  
Examiner, Art Unit 2626  
Patent Examiner  
AU 2626  
(571)-270-1847  
[Michael.Colucci@uspto.gov](mailto:Michael.Colucci@uspto.gov)

/Richemond Dorvil/  
Supervisory Patent Examiner, Art Unit 2626